Atmospheric Pollution Research 6 (2015) 540-549

Atmospheric Pollution Research

www.atmospolres.com

Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales

Ana Russo [1], **Pedro G. Lind** [2,3], **Frank Raischel** [1], **Ricardo Trigo** [1], **Manuel Mendes** [4]

[1] *Instituto Dom Luiz, Faculdade de Ciências da Universidade de Lisboa, Campo Grande Edifício C8, Piso 3, 1749–016 Lisboa, Portugal*
[2] *TWIST – Turbulence, Wind energy and Stochastics, Institute of Physics, Carl–von–Ossietzky University of Oldenburg, DE–26111 Oldenburg, Germany*
[3] *ForWind – Center for Wind Energy Research, Institute of Physics, Carl–von–Ossietzky University of Oldenburg, DE–26111 Oldenburg, Germany*
[4] *Instituto Português do Mar e da Atmosfera, Rua C–Aeroporto,1749–077 Lisbon, Portugal*

ABSTRACT

We present a simple neural network and data pre–selection framework, discriminating the most essential input data for accurately forecasting the concentrations of PM_{10}, based on observations for the years between 2002 and 2006 in the metropolitan region of Lisbon, Portugal. Starting from a broad panoply of different data sets collected at several air quality and meteorological stations, a forward stepwise regression procedure is applied enabling to automatically identify the most important variables for predicting the pollutant and also to rank them in order of importance. The importance of this variable ranking is discussed, showing that it is very sensitive to the urban location where measurements are obtained. Additionally, the importance of Circulation Weather Types is highlighted, characterizing synoptic scale circulation patterns and the concentration of pollutants. We then quantify the performance of linear and non–linear neural network models when applied to PM_{10} concentrations. In the light of contradictory results of previous studies, our results show no clear superiority for the case studied of non–linear models over linear models. While all models show similar predictive performances, we find important differences in false alarm rates and demonstrate the importance of removing weekly cycles from input variables.

Keywords: *Pollution, PM_{10}, forward stepwise regression, circulation weather types, neural networks*

Corresponding Author:
Ana Russo
☎ : +351-217-500-818
🖷 : +351-217-500-807
✉ : acrusso@fc.ul.pt

Article History:
Received: 30 September 2014
Revised: 15 December 2014
Accepted: 15 December 2014

doi: 10.5094/APR.2015.060

1. Introduction

Air pollution is a global threat to public health and to the environment, particularly in urban areas (Kolehmainen et al., 2001; EEA, 2013). Urban air pollution is a complex mixture of toxic components, which may induce acute and chronic responses from sensitive groups, (Kolehmainen et al., 2001; Wong et al., 2002; Diaz et al., 2004). Therefore, forecasting air pollution concentrations in urban locations emerges as a priority for guaranteeing life and environmental quality (Kolehmainen et al., 2001; EEA, 2013).

Modeling air pollution allows describing the causal relationship between emissions, meteorology, atmospheric concentrations, deposition, and other factors, including the determination of the effectiveness of remediation strategies, and the simulation of future scenarios. Despite of the above mentioned advantages of pollution modeling, the choice for a certain modeling approach should be done with some parsimony. Particularly, the time–lag in which air pollution prediction is performed should allow effective alert procedures in urban centers.

Different methodologies have been applied to characterize and forecast the dispersion of air pollutants, from the most simple approaches, such as box models (Middleton, 1998), or persistence and regression models (Shi and Harrison, 1997), to the most complex dynamical model systems, such as CHIMERE (Monteiro et al., 2005), or the CMAQ–Community Multiscale Air Quality Model (Luecken et al., 2006; Arasa et al., 2010).

Simpler models are often used as they can provide a fast overview. However, they rely on significant simplifying assumptions and usually do not describe the complex processes and interactions that control the transport and chemical behavior of pollutants in the atmosphere (Luecken et al., 2006).

In the last decades, significant progress has been made in air–quality dispersion models (Arasa et al., 2010). However, being highly non–linear, they require large amounts of accurate input data and are computationally expensive (Dutot et al., 2007; Elangasinghe et al., 2014).

Statistical models, such as Artificial Neural Networks (NN), have been shown to constitute a promising alternative to deterministic models (Yi and Prybutok, 1996; Cobourn et al., 2000; Gardner and Dorling, 2000a; Hooyberghs et al., 2005; Dutot et al., 2007; Papanastasiou et al., 2007; Lal and Tripathy, 2012; Nejadkoorki and Baroutian, 2012; Elangasinghe et al., 2014). These models are often regarded as a good compromise between simplicity and effectiveness, being capable of modeling the effect of non–linearities and fluctuations.

Although NN models may involve greater uncertainty, the input data requirements are less strict. Several NN models have been tested comparing the potential of different approaches when applied to different pollutants and prediction time lags (Yi and Prybutok, 1996; Gardner and Dorling, 2000a; Kukkonen et al., 2003; Hooyberghs et al., 2005). Other authors have proven better forecasting results of NN over multiple linear regression (MLR) (Kukkonen et al., 2003; Agirre–Basurko et al., 2006). More recently, Russo et al. (2013) showed that, combining NN models and stochastic data analysis, allows diminishing the requirement of large training data sets often appearing when constructing a NN model.

Despite these improvements, forecasting NN models still present some caveats that need to be properly addressed (Lal and Tripathy, 2012). The construction of the best NN structure and the choice of input parameters constitutes a challenge (Chaloulakou et al., 2003; Hooyberghs et al., 2005; Perez and Reyes, 2006; Lal and Tripathy, 2012), as any set of input data can be fed into any NN architecture for training and evaluation, but not all possible combinations can be realistically tested. Comrie (1997) and Cobourn et al. (2000) have performed comparison studies between NN and regression models to forecast ozone concentrations, both showing that NN outcomes are only equal or slightly better than regression. In contrast, Gardner and Dorling (2000b) showed that there is a significant increase in performance when using non–linear models. For PM_{10}, the results are different to some extent, and it is possible to find in the literature different applications where NN can perform well, depending on input parameters (Chaloulakou et al., 2003; Perez and Reyes, 2006; Nejadkoorki and Baroutian, 2012). Comparison statistics between linear and nonlinear models presented by Chaloulakou et al. (2003) and Perez and Reyes (2006) indicate that the NN approach has an edge over linear models, expressed both in terms of prediction error and of episodic prediction ability, demonstrating that NN models, if properly trained and formed, can provide adequate solutions to particulate pollution prognostic demands. Thus, a good choice of input variables appears to be very important (Chaloulakou et al., 2003; Perez and Reyes, 2006; Hooyberghs et al., 2005) and should be performed with parsimony. Even though several studies revealed that certain weather parameters are relevant to model air pollutant concentrations (e.g. temperature, wind speed and direction, humidity) (Hooyberghs et al., 2005; Demuzere et al., 2009), the majority of the research focused on individual meteorological variables and non–automated procedures of variables' selection. Moreover, several studies have been published establishing important links between synoptic scale circulation patterns, usually named Circulation Weather Types (CWT), and air pollution (Dayan and Levy, 2002; Demuzere et al., 2009; Saavedra et al., 2012; Russo et al., 2014), relating a particular air mass to dispersion conditions and also to the mesoscale and local meteorological behavior (Dayan and Levy, 2005). Nevertheless, to the best of our knowledge, there are no studies in the literature focusing on the application over the Iberian Peninsula of objective automatic classification procedures of CWT as a predictor for air quality modeling.

In this paper, we address the issues previously mentioned, (1) aiming at developing daily forecast through the application of a circulation–to–environment approach based on the analysis of links between meteorological parameters, CWT and daily air quality measurements, and (2) introducing a simple framework for automatically ranking the set of variables used as input variables for training the NN model. To systematically develop a better air quality model, we apply both linear and non–linear NN models to predict PM_{10} daily average concentrations within the greater urban area of Lisbon, Portugal, based on historical air pollution and weather information. We choose to address only PM_{10}, that corresponds to inhalable particulate matter sized 10 µm or less, as this pollutant poses a major health risk (Stedman et al., 2002). Although pollutants' emissions in Europe have decreased over the last two decades, this did not lead to a corresponding reduction of concentrations of PM_{10} throughout Europe (EEA, 2011). Evidence has accumulated during the last years that there is a direct association between daily variations in the concentrations of airborne particles and a range of health indicators (Stedman et al., 2002; Wong et al., 2002; Diaz et al., 2004).

Despite the mitigating impact of the nearby Atlantic Ocean on the effects of aerosols and pollution (Almeida et al., 2013), Lisbon has been affected by several high pollution episodes in the last two decades, exceeding repeatedly the legal limits imposed for PM_{10} (APA, 2008; Russo et al., 2014). Those episodes are often related to the occurrence of synoptic patterns with an eastern component which results in an eastern/southeastern flow and advection of dryer continental air (Russo et al., 2014). Therefore, a good PM_{10} prediction procedure with a sufficiently large time–lag is needed to prevent the occurrence of exceeding concentrations.

The methodological approach here presented is very straightforward in terms of operational implementation and has low computational costs and thus can be relevant for daily surveillance and alert systems in the Lisbon area.

2. Data

2.1. Target data

We consider daily values of PM_{10} concentrations measured by twelve monitoring stations in the agglomeration of Lisbon (Figure 1), between 2002 and 2006, which record the atmospheric concentrations of major pollutants, such as gases (e.g. NO_2, NO and CO) and PM_{10}. This network is complemented by three meteorological monitoring stations, located near the stations of Avenida da Liberdade (AL), Lavradio (L) and Olivais (O).

A preliminary data analysis showed that it is difficult to identify a clear cycle in PM_{10}, cf. Figure S1 of the Supporting Material (SM). However, when analyzing Figure S1a, it is possible to identify higher values during winter and summer months and lower ones during autumn and spring. Nevertheless, the cyclic behavior is not as noticeable as it usually occurs with O_3 and NO_2, cf. Figure S1 of the SM.

Daily legal limits were often exceeded during the 2002–2006 period in all the monitoring stations (APA, 2008), but the number of days with exceeding values is especially impressive for AL and E (Entrecampos) stations. It is worth mentioning that, in both stations two types of exceedances occurred, as the daily legal limit (50 µg/m^3) was exceeded, but also the number of times that the daily limit can be exceeded per year (35 exceedances/year) was also surpassed (APA, 2008).

Thus, the target of the present work is to predict PM_{10} on day t+1 on each monitoring station based on measurements on day t of several input variables (Section 2.2.).

2.2. Input data for NN training

The 15 variables that are available as NN input data sets are shown in Table 1. Additionally to the pollutant's concentration measured on the previous day and at 00:00 UTC (Universal Time Coordinated), several available meteorological variables measured in the 3 monitoring stations were considered.

In order to include information regarding the atmospheric stability and circulation, which is an important factor for the accumulation of pollutants near the surface, two other variables were considered, namely the boundary layer and the daily CWT. Three boundary layer height (BLH) fields were retrieved from the ECMWF 40–years reanalysis (ECMWF, 2013) for the years 2002–2006: the 03:00 UTC (BLH5), 09:00 (BLH7) and 21:00 UTC (BLH11). The BLH varies along the day, and the anti–phase diurnal variations of PM mass concentrations and BLH indicate that the BLH is one of the important factors affecting air quality (Du et al., 2013). Thus, we decided to use 3 measures of the BLH, one during night time (BLH5), one during peak traffic hours (BLH7) and one after the normal work day ends (BLH11). The daily CWT classification was determined based on the Trigo and DaCamara (2000) approach, and 10 classes were retained (see Section 3.1), eight of which are directional (NE, E, SE, S, SW, W, NW, and N) and two are dominated by the shear vorticity (cyclonic C or anticyclonic A).

Table 1. *Input parameters used for training the NN. Daily CWT and daily mean wind direction are categorical variables*

Variables (Lag=1 day)	Units
Daily mean concentration of NO_2, NO, CO, PM_{10}	μg/m^3
Daily maximum concentration of PM_{10} (PM_{10} m)	μg/m^3
Concentration of PM_{10} at 00:00 UTC	μg/m^3
Daily CWT	
Boundary layers heights:	
BLH5 (03:00 UTC)	m
BLH7 (09:00 UTC)	m
BLH11 (21:00 UTC)	m
Daily maximum temperature (T_{max})	°C
Daily mean wind direction (V_d)	
Daily mean wind intensity (V_i)	m/s
Daily mean humidity (Hum) and radiance (Rad)	%, KJ/m^2

Based on the available five–year datasets (between 1/1/2002 and 31/12/2006), we constructed a collection of records, consisting of the input vector, which included the meteorological variables, air pollutant concentrations, and the corresponding target PM_{10}. The first four years were used to construct the models and the year 2006 was used for independent evaluation.

The different datasets collected from several monitoring stations in Lisbon were used to train the NN models. The predictions will allow a better definition of air pollution episode alerts with spatial variability.

3. Methods

3.1. Circulation–to–environmental approach

The concentration of pollutants in the atmosphere is linked to the occurrence of certain synoptic weather conditions (Demuzere et al., 2009) and to the regional wind flow pattern induced by mesoscale meteorological processes such as land–sea breezes (Dayan and Levy, 2005). Prevailing CWTs often dictate the long–range transport, linking a particular air mass to dispersion conditions and also to the mesoscale meteorological configuration that controls the regional transport of air pollution (Dayan and Levy, 2005). These prevailing CWT classifications have witnessed a growing interest by the research community during the last two decades (e.g. Trigo and DaCamara, 2000; Demuzere et al., 2009; Russo et al., 2014).

The majority of CWT classification procedures are based on the application of statistical selection rules (e.g. cluster analysis, regression trees), but can also be based on the determination of physical parameters regarding the prevailing atmospheric circulation pattern. Furthermore, CWTs are generally specific to a given region and therefore it is possible to evaluate the relative importance of each CWT in pollutant concentration forecasts.

We use here the CWT objective classification which has successfully been applied to Portugal mainland by Trigo and DaCamara (2000) who linked different CWTs to precipitation.

Values for daily mean sea level pressure (SLP), relative humidity and temperature and geopotential height at the 1 000 hPa level values were extracted from ERA–Interim Reanalyses dataset (Dee et al., 2011) on a 1° latitude by 1° longitude grid for Portugal (40W–30E, 20–70N). The period between 1981 and 2010 was used to derive a 30 year climatology that included the air quality period under analysis (2002–2006). Based on the large–scale fields, prevailing CWT at regional scale were determined and were then considered as an input variable.

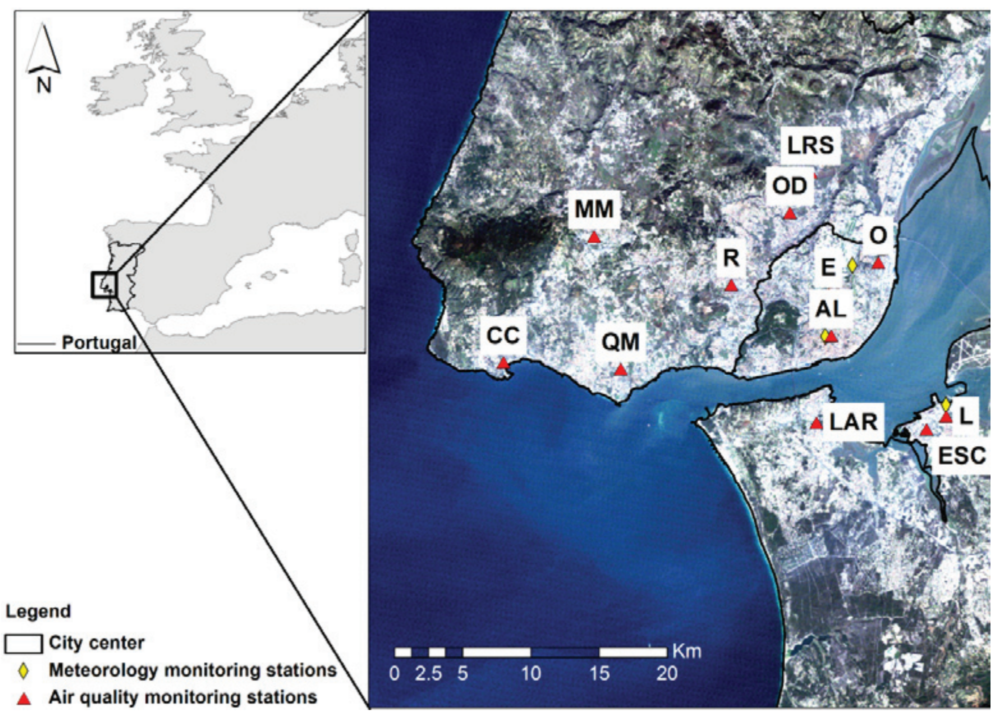

Figure 1. *Air quality (triangles) and meteorological (diamonds) monitoring stations in the region of Lisbon (Portugal). Each data set is extracted within the period between 2002 and 2006, with a daily sampling frequency.*

3.2. Choice of predictors

A crucial step in the development of a forecast model is the choice of input parameters, the predictors (Hooyberghs et al., 2005). Usually, a number of statistical methods can be applied in order to choose the most appropriate set of predictors/inputs. Important methods in this scope include: stepwise regression (SR), principal component analysis (PCA), cluster analysis and ARIMA (Wilks, 2006). These methods are pre–processing procedures, which allow reducing the number of input variables into the models, thus considerably diminishing redundant information, instabilities and over–fitting.

Here, the selection of variables was made independently for each monitoring station through a forward stepwise regression (FSR), from which the best time lag for each input variable was also determined. During this procedure, which starts with the variable most correlated with the target, additional variables are added which, together with the previously selected variables, most accurately predict the target (Wilks, 2006). The procedure stops when any new variable does not significantly reduce the prediction error. Significance is measured by a partial *F*–test applied at 5% (Wong et al., 2002; Wilks, 2006).

The use of an automated procedure prior to NN modeling also allows improving the quality and robustness of pollutant concentration forecasts, which are crucial properties when linking the forecast to alert systems.

3.3. The neural network framework

Neural networks are mathematical models inspired by the biological nervous system (Gardner and Dorling, 1998; Cobourn et al., 2000; Agirre–Basurko et al., 2006). One of the most common examples of architectures used is the multilayer perceptron (Haykin, 1999; Agirre–Basurko et al., 2006), where the artificial neurons can be organized following different types of architectures, composing a certain number of levels (Figure 2) (Haykin, 1999; Agirre–Basurko et al., 2006). In the zero level one has the set of independent variables, X_i (i=1,...,p), and a number of connections with a weight ω_{ij} (i=1,…,p and j=1,…, number of hidden neurons), joining the variables X_i to neurons in the next level (Gardner and Dorling, 1998; Trigo and Palutikof, 1999). In the first level ("input layer" in Figure 2), each neuron computes a linear combination of the weighted inputs $\omega_{ij}X_i$, including a bias term b_j: $Y_i=\Sigma_i\omega_{ij}X_i+b_i$. This sum is transformed using a linear or non–linear activation function, W_i=f(Y_i). The neuron activation function can be any smooth function. Some of the commonly used functions are the linear function, the Gaussian function, the sigmoid function, the hyperbolic function, the inverse tangent, among others (Haykin, 1999). The weights can initially be chosen randomly, and are then properly adjusted during the training of the NN as described below. The bias term is included in order to allow the activation functions to be offset from zero and it can be set randomly or set to a desired value, such as a dummy input with a magnitude equal to one.

The output W_j obtained at the previous level is then passed as an input to other nodes in the following layer, usually named hidden layer. This procedure is performed repeatedly to better tune the weights until a certain accuracy threshold between the produced output and the target variable (empirical data) is reached. The accuracy threshold between the output and the target variable is imposed in the beginning of the iterative procedure and usually is based on an early–stopping procedure in order to prevent overtraining (Haykin, 1999). Two of the simplest ways to perform the early–stopping were applied: (1) to limit the number of iterations to a predetermined value or to choose an acceptable error level for the problem; and (2) apply a threshold to the error between prediction and observations (Haykin, 1999). It is possible to use several hidden levels, successively. However, it is often advantageous to minimize the number of hidden nodes and layers, in order to improve the generalization capabilities of the model and also to avoid over–fitting (Gardner and Dorling, 1998).

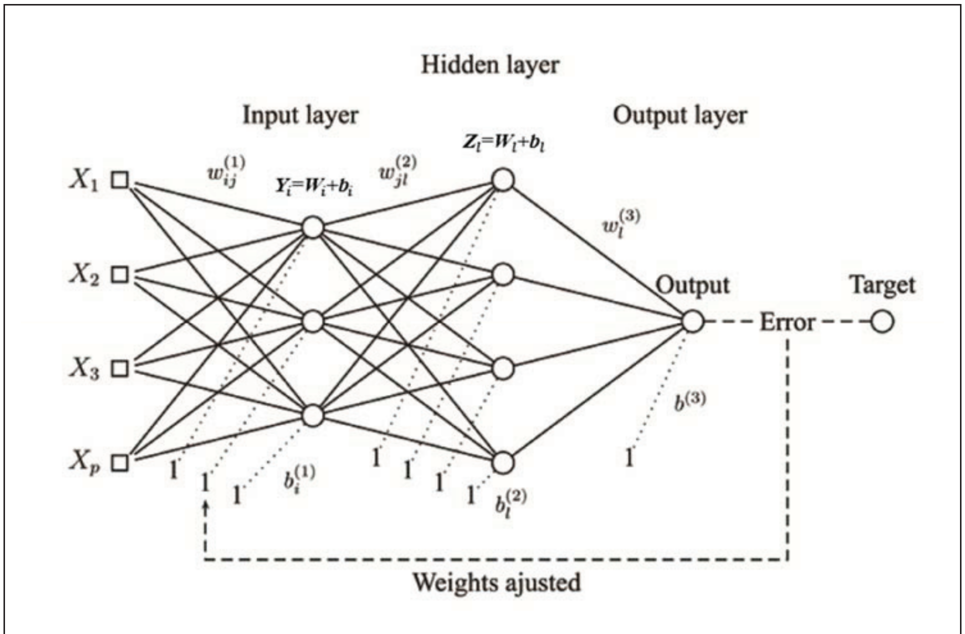

Figure 2. *Illustration of a feed–forward artificial neural network model with three layers. Input variables X_i=1,...,p can be perceived as "neurons", ω_{ij} represent the weights associated to each neuron and b are the bias vectors which combined will produce an output within certain error limits. (Adapted from Trigo and Palutikof, 1999).*

There are several training procedures for estimating the weights and associate input and output. Here, we use a modified version of the back–propagation (BP), one of the most popular and common training procedures (e.g. Haykin, 1999; Trigo and Palutikof, 1999). BP has drawbacks, as any other procedure, being often slow to converge and the final weights may be trapped in local minima over the highly complex error surface (Haykin, 1999; Trigo and Palutikof, 1999). To overcome this shortcoming, numerically optimized techniques have been developed, such as the Levenberg–Marquardt method (LM) which ise based on an approximation of the Gauss–Newton method. Thus, we have used the LM method for the nonlinear NN models as it holds the advantage of converging faster and with a higher robustness than most of the basic BP schemes (Trigo and Palutikof, 1999) because it avoids computing second–order derivatives. For the linear approach we have used the Widrow-Hoff rule.

Unlike the standard BP algorithm that can be trapped in local minima, the Widrow–Hoff rule will give a unique solution corresponding to the absolute minimum value of the error surface (Haykin, 1999). For this reason, we chose to use the Widrow–Hoff rule for the much simpler convergence of the linear approach.

The linear NN model is composed of a single one–layer NN structure with just one neuron, which employs a linear activation function and behaves exactly like a linear model, producing the same results as a linear regression model (Weisberg, 1985). Following other previous applications, this type of NN structures constitute the baseline against which the performance of the non–linear model will then be compared. The non–linear NN models used here are based on a feed–forward configuration of the multilayer perceptron that has been used by several authors (Haykin 1999; Hooyberghs et al., 2005; Papanastasiou et al., 2007; Nejadkoorki and Baroutian, 2012). For the linear model, a perceptron with a linear activation function was used, while for the non–linear models, the log–sigmoid function was used, except for the single node in the output layer, for which we consider a linear transfer function.

For the sake of simplicity we will refer from this point forward to the linear model as MLR and to the non–linear model as NN solely.

3.4. Application of the NN framework

Since the factors mainly contributing to air pollution concentration are connected with source activity that can present periodic variations, it is normal to expect periodic components in air quality time series (Kolehmainen et al., 2001). Hence, following a similar approach to the study presented by Kolehmainen et al. (2001), two modeling approaches are possible. One is to model the original data signal. Another is to model the residual component after the removal of a periodic component from the original signal. Here, we address the relative importance of the weekly periodic component, which is mainly affected by traffic and weekly business and industrial fluctuations. The forecasting capabilities of the two different approaches are compared.

Thus, the NN framework is applied to our data set in the following way. Consider an attribute $Z(x,t+\Delta t)$, symbolizing the concentration of PM_{10}, measured at a spatial location x at day t, which yields a daily series of the pollutant's concentration at each monitoring station. Δt represents the prediction's temporal lag, which is usually hourly or daily (Gardner and Dorling, 1998; Haykin, 1999; Cobourn et al., 2000; Hooyberghs et al., 2005; Agirre–Basurko et al., 2006; Papanastasiou et al., 2007). One considers then its decomposition into a periodic component $M(x,t+\Delta t)$ and a residual $R(x,t+\Delta t)$, yielding $Z(x,t+\Delta t)=M(x,t+\Delta t)+R(x,t+\Delta t)$.

In particular we consider the periodic components M_7($x,t+\Delta t$), which is determined respectively by a 7 days moving average. Likewise, we take also the respective residuals R_7($x,t+\Delta t$) obtained from the removal of the correspondent periodic component.

Moreover, the variables were transformed prior to the processing. This was carried out by determining the maximum and minimum values over the whole data period and calculating normalized variables using the following formula:

$$X_{norm} = \frac{X - \frac{Max(X) + Min(X)}{2}}{\frac{Max(X) - Min(X)}{2}} \tag{1}$$

We then apply the linear and non–linear models, i.e. MLR and NN models, to each monitoring station in order to model both the complete signal, hereafter called TOT approach, and to model the residual components, hereafter called RES approach. The forecasting capabilities of the different approaches are compared in order to assess the potential improvement using non–linear NN in air quality modeling.

We tested a large number of architectures, each one with a given number of hidden layers. The use of two layers was verified to be sufficient, since a superior number of layers do not improve the output.

In both linear and nonlinear cases, a cross–validation is applied to the 4 years available for calibration and validation purposes, i.e. each time three years are used for construction of the neural network model and the remaining year is retained for validation. Thus, the first run is performed using data for 2002–2004 to train the model, whereas data from 2005 is used for validation purposes. In the second run, data for 2003–2005 are used for training and data for 2002 for validating, and so on. With such cross–validation procedures (Wilks, 2006), it is possible to account for the risk of over– or underfitting. Moreover, in this way, one is able to ascertain if the models are stable and if they are capable of generalizing correctly in forecast mode. After the calibration and validation procedure with historical data (2002–2005), the models are used to produce forecasts for the daily average of PM_{10} concentration, during a period of one year. For this purpose an independent one–year sample, the year 2006, is left out in order to be used for evaluation of models performance (Section 3.5) during the individual daily average predictions. Finally, the forecasts are then compared with the actual observed pollutant values at the monitoring stations.

3.5. Performance indicators

Rigorous quantitative measures are required to perform models' evaluation. Thus, in order to evaluate the efficiency and performance of the developed models three continuous performance indicators are used. The simplest measure is the Pearson correlation coefficient (PC):

$$PC = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(o_i - \bar{o})}{\sqrt{\sum_{i=1}^{N}(o_i - \bar{o})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} = \sqrt{R^2} \tag{2}$$

where, y_i denotes the respective model forecast at time i while o_i denotes the real observed values at time i, and $\bar{y}$ and $\bar{o}$ are the corresponding average values. For the case of a linear regression, PC corresponds to the square root of the Coefficient of Determination, R^2 (Wilks, 2006).

A quantity similar to PC, also related to correlation between series, is the root mean square error (RMSE) given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2} \tag{3}$$

Considering that correlation coefficients are not robust to deviations from linearity, its exclusive use to evaluate the quality of a model can lead to misleading results (Wilks, 2006). Therefore, we consider these quantities combined with other properties which present different abilities for accessing important aspects of the data such as outliers and average values.

The skill against persistence, SSp, which is interpreted as the percentage of improvement that our model can provide when compared with the persistence model (Trigo and Palutikof, 1999; Wilks, 2006), i.e. the model that yields the observed value of yesterday as the forecast for today. The SSp is also used as a measure of the relative accuracy of the model. The score is quantitatively defined as:

$$SS_p = \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - o_i)^2 - \frac{1}{N-1}\sum_{i=1}^{N-1}(o_{i+1} - o_i)^2}{\frac{1}{N-1}\sum_{i=1}^{N-1}(o_{i+1} + o_i)^2} * 100 \quad (4)$$

Both linear and non–linear models will be compared with this persistence model, which is the simplest way of producing a forecast and assumes that the conditions at the time of the forecast will not change. Due to a certain level of memory that characterizes air pollutants, persistence corresponds to a benchmark model considerably more difficult to beat than climatology or randomness (Demuzere et al., 2009).

Additionally, four categorical measures are also considered, to ascertain if the models are able to predict exceedances (Wilks, 2006). Traditional categorical metrics used in model evaluations assess the model's ability to predict an exceedance which is defined by a fixed threshold. These metrics are defined by sets of observational forecasts that are paired together. Here, we used the false alarm rate (F), i.e. the proportion of non–occurrences incorrectly forecasted, and the proportion of correctness (PCS), i.e. the proportion of events properly forecasted. Both categorical measures, F and PCS, are applied against binary time series obtained with thresholds for poor air quality limit values (PM_{10}: 50 $\mu g/m^3$). A detailed description of these two categorical measures can be found in Jolliffe and Stephenson (2003).

One should however notice that there is now considerable evidence that daily hospital admissions for cardiorespiratory diseases are linked to levels of PM_{10} not only on the same, but also on previous days (Wong et al., 2002) and that association is positive for values lower than the legal thresholds. Thus, two additional categorical measures were introduced in order to assess if the models are able to perform correctly for a new threshold that corresponds to 50% of the legal limit value (F50 and PCS50).

4. Results and Discussion

4.1. Selection of input variables

We first consider all 15 potential predictors for PM_{10} (Table 1). The use of the FSR has reduced the complexity by retaining substantially less variables, namely only those marked in Table 2. The numbers given in Table 2 correspond to the rank of relevance of each variable for that specific model, i.e., monitoring station and approach.

Table 2. *Available and chosen predictors by FSR for each monitoring station for both the TOT and RES–7 approaches. The numbers correspond to the rank of relevance of each variable for that specific model*

		Stations											
		E	O	AL	L	ESC	R	LAR	LRS	CC	QM	MM	OD
PM_{10}	TOT	5	2	2	2	3	2	2	6	2	3	4	7
	RES–7							4		9			
PM_{10} at 00:00 UTC	TOT	1	1	1	1	1	1	1	1	1	1	1	1
	RES–7	1	1	1	1	1	1	1	1	1	1	1	1
CO	TOT					5	9	7	9			5	
	RES–7	5				4	3		9	7	7	6	
NO_2	TOT			7	8	6			10				3
	RES–7	3	4	2		2			8	8	6	8	3
NO	TOT			8						4	5		4
	RES–7										5		5
PM_{10} m	TOT	6	7				8		7		7	7	6
	RES–7								7				4
V_d	TOT	3	4	3	4		4	4	3	3	4	3	8
	RES–7	4	3	3	3		4	3	3	2	2	3	
V_i	TOT					7							
	RES–7												
Rad	TOT				5						9		
	RES–7												
Hum	TOT	2	3	4	7	2	3	3	5				
	RES–7	2	2		2		2	2	2	6		2	2
T_{max}	TOT	4	5	6	3		6	5	2	5	2	2	2
	RES–7				6			6	5	3	3	5	7
CWT	TOT												
	RES–7	7			5			7		5	9	9	8
BLH5	TOT	7	8	5			7		8	6	6		
	RES–7									4	4		
BLH7	TOT					4							5
	RES–7		5			3	5		4			4	6
BLH11	TOT	8	6		6		5	6	4		8	6	
	RES–7	6			4			5	6		8	7	

We also found that adding time lags superior to one day does not provide relevant additional information. Therefore, only the one–day time lag for both meteorological and air quality variables are taken into account in the subsequent analysis. Our analysis further revealed that the most significant variable in predicting PM_{10} for all the monitoring stations is the 00:00 UTC PM_{10} concentration. The introduction of several PM_{10} measures as potential predictors was tested bearing in mind that their introduction raises issues of multicollinearity. As a first attempt, we have tried to model average PM_{10} concentrations solely based on the average value observed on the previous day in association to other pollutants and meteorological variables, i.e., without using the maximum and the hourly concentration. The model performances were poorer and most of all, the models were not able to capture large changes between two consecutive days neither peak events. After the introduction of those two other PM_{10} measures, the performance of the models increased indicating that the models become more robust. In order to avoid multicolinearity we have applied a stringent validation procedure.

Based on the analysis of Table 2, we also noticed that the most significant variables in predicting PM_{10} are pollutants related to road traffic emissions and meteorological variables related to atmospheric stability (i.e., BLH and CWT). Moreover, other variables that were retained for the majority of the stations under the TOT approach are the previous day average and maximum PM_{10} concentrations, the previous day average values of NO_2, NO and CO concentrations, the maximum temperature, wind direction, humidity and BLH. The pollutant variables retained for the majority of the stations under the RES–7 approach are the previous day average values of NO_2 and CO concentrations. Particularly for the RES–7 approach, and apart from the pollutants, the most significant meteorological variables in predicting PM_{10} are the maximum temperature, the daily wind direction, humidity, CWT and BLH7 and BLH11. The two referred BLH correspond to the periods of the day when the traffic is more intense in the city of Lisbon. It should be noted that, although the previous day concentration PM_{10} has a very high rank for the prediction of the TOT series, it almost has no impact on the prediction of the RES–7 series; this implies that in the first case it merely reflects the periodic aspects of the data.

Additionally, it is worth stressing that, the RES–7 approach includes the CWT classification as one of the most important predictors in the majority of the monitoring stations. The dependence on the wind, relative humidity, cloud cover and BLH which is highlighted on Table 2 was also shown in previous works (Hooyberghs et al., 2005; Demuzere et al., 2009). Moreover, the NO_2 and CO dependence in some stations is present due to road traffic influence, as road traffic behaves as a local source of PM_{10} (Demuzere et al., 2009).

Kukkonen et al. (2003) showed that the inclusion of meteorological variables for the day of prognosis, i.e., predictions of the meteorological variables for the next day, improves the performance of NN models and that linear models perform significantly worse in this situation. However, we consider that these variables might unnecessarily increase the error associated with the prediction and choose not to include them at this stage.

Another possible way of reducing the complexity by retaining substantially less variables or information is through the application of a hybrid technique which combines regression analysis techniques and feedforward backpropagation with PCA. That type of approaches main aim is to reduce the complexity of the model, reduce collinearity of the models and determine the relevant independent variables to predict future PM_{10} concentrations. However, it also presents some caveats as the distribution function of the input variables used. Air pollutants usually have very skewed distribution functions and the application of a PCA assumes a Gaussian behavior. Nevertheless, the comparison between performances regarding the hybrid approach is out of the scope of this paper.

4.2. Comparison of methods

The validation tests presented here are based on the use of MLR and NN models in which all the retained predictor variables are incorporated according to the framework in Table 2. Validation results obtained with the MLR and NN models are shown in Table 3. The numbers of hidden neurons used are identified by the numeric index after NN, i.e., NN2 refers to a NN model with 2 neurons in the hidden layer. The choice of the number of hidden units was made iteratively. There are four main conclusions to be drawn from Table 3:

1. All the models perform substantially better than persistence with SSp scores above 45%.
2. The proportion of correctness (PCS and PCS50) is quite high which indicates that the models are robust and able to correctly predict medium values and also higher value events.
3. The false alarm rate (F) is significantly low for high values (i.e., for values above the legal threshold), which indicates that only a low percentage of non–observed pollution episodes are wrongly forecasted.
4. Overall, weekly residuals (RES–7) models outperform the TOT models in most evaluated measures. Removing the weekly cycle appears to be a promising approach compared to the complete signal model (TOT).
5. The RES–MLR model performs approximately the same as RES–NN2 and RES–NN3, and considerably better than TOT–MLR models.

Although the MLR presents better performances for some of the indicators than the NN models, the MLR is subject to the collinearity problem. RES–7–NN2 does not suffer from collinearity and some of the performance measures are more robust in some way, but has the disadvantage of having convergence problems. Under such type of situation, the simplest model should be chosen (Demuzere et al., 2009) and henceforth, we will restrict the remaining analysis to the MLR approach.

From the operational point of view, the effectiveness of a prediction model should be judged according to its ability to forecast properly in order to be able to alert the population and the competent health authorities. However, the forecast models are known a priori to be imperfect, thus the alert threshold must be set below the critical level objectively identified, in order to allow for a margin of safety (Cobourn et al., 2000).

Table 3. *Average performance indicators obtained for the PM_{10} calibration/validation process, including the Pearson correlation coefficient (PC), the skill against persistence [SSp (%)], the root mean square error [RMSE (μg/m^3)], the false alarm rate [F (%)], the proportion of correctness [PCS (%)] , the 50% false alarm rate [F50 (%)], and the 50% proportion of correctness [PCS50 (%)]. Each average performance indicator was determined based on the indicators of all the monitoring stations*

Model	PC	SSp	RMSE	F	PCS	F50	PCS50
TOT–MLR	0.75	45.00	12.85	6	88	50	80
RES–7–MLR	0.81	54.41	11.69	12	86	62	89
RES–7–NN2	0.81	54.30	11.69	11	85	64	90
RES–7–NN3	0.81	54.20	11.69	11	85	63	90

Still, the performance indicators presented here are superior to those obtained by Demuzere et al. (2009) for the Netherlands and by Nejadkoorki and Baroutian (2012) for Iran, and are consistent with the results presented by Hooybergs et al. (2005) for Belgium. Here we attain similar results, 0.75<PC<0.81, by incorporating meteorological variables. Moreover, checking the performance results, one observes a tendency for higher performance for the independent validation, which is due to the favorable characteristics of year 2006, as we explain in the next section.

4.3. Forecast: Independent validation

The forecasts retrieved by the MLR models were compared with the actual observed pollutants values for the year 2006 for all the monitoring stations available. The scatter plots and correlation coefficients between observed and modeled values were computed for all monitoring stations.

Figure 3 presents the aggregated scatter plots and correlation coefficient for all monitoring stations. The results for the independent sample show a very high average correlation (PC>0.84) between the predicted and observed values.

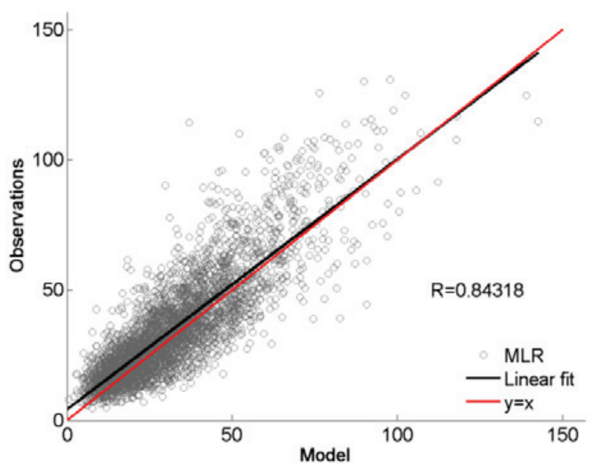

Figure 3. *Scatter plots of MLR results versus actual observed PM_{10} values for all monitoring stations and for the year 2006.*

In Table 4 the correlation coefficients for each individual monitoring station for the calibration/validation period (2002–2005) and for the one–year independent evaluation period (2006) are presented. These results show that MLR model generalizes well for independent data and for each monitoring station.

In general, MLR techniques are known to underestimate peak levels. Interestingly, although the MLR model is built using the calibration dataset only, we can observe an increase in accuracy for the majority of the stations when in forecast mode. This may be explained by the particular characteristics of the historical data used to construct the models, namely the years 2003 and 2005.

The PM_{10} data sets used on this work comprehend the years from 2002 to 2006 in Lisbon. For this location, the years of 2003 and 2005 were particularly outstanding relatively to weather conditions, namely an exceptional heat wave that struck the entire western Europe, in 2003 (Trigo et al., 2006) and one of the most severe droughts of the 20th century occurred in 2005 (Garcia–Herrera et al., 2007).

Moreover, air pollution is strongly influenced by shifts in the weather. Changes in the temperature, humidity and wind indeed induce changes in the transport, dispersion, and transformation of air pollutants at multiple scales (Dias et al., 2012). Therefore, using all the years as individual calibration/validation samples, yields quite disparate skill values on one hand with an average that is significantly below the skill against persistence obtained when using these anomalous years for independent validation of 2006.

Table 4*. Correlation coefficients between observed and modeled PM_{10} concentrations for each station considered and for the calibration/ validation period (2002–2005) and for the independent forecast year (2006)*

Station	2002–2005	**2006**	Δ%
E	0.83	0.78	–5
O	0.79	0.86	7
AL	0.81	0.82	1
L	0.83	0.86	3
ESC	0.80	0.83	3
R	0.79	0.87	8
LAR	0.85	0.87	2
LRS	0.83	0.87	4
CC	0.75	0.78	3
QM	0.83	0.86	3
MM	0.82	0.86	4
OD	0.85	0.82	–3

5. Conclusions

In this paper we introduce a framework consisting in a pre–selection procedure of predictors which are then used as input data to train NN model.

Our framework enables to rank all given variables and then select the highly ranked variables as predictors, which were chosen for each monitoring station separately. To rank the variables a forward stepwise regression was used. We found that the most significant variables in predicting PM_{10} are pollutants related to road traffic emissions and meteorological variables related to atmospheric stability. Particularly for the RES–7 approach, the most significant variables in predicting PM_{10} are, in descending order of importance, the 00:00 UTC PM_{10} concentration, the previous day average values of NO_2 and CO concentrations, the maximum temperature, the daily wind direction, humidity, CWT and BLH7 and BLH11. These results emphasize the importance of meteorological variables and of the circulation–to–environment approach to air quality forecast.

In particular, we found that for forecasting PM_{10} in Lisbon, CTW should be taken as input data, though its rank is not particularly high compared with other meteorological data. However, we point out that the ranking of predictors varies considerably from one station to another, since it reflects the diversity of geographical and urban features, such as traffic, industries, and distance to the coast. Therefore, a forthcoming approach to urban pollution would be to apply such procedure to a panoply of different pollutants and ascertain which ones are more sensitive to synoptic scale circulation and meteorological constraints. Another issue to be addressed in a forthcoming study is the interaction between stations.

In order to assess the importance of the periodic and residual components present in pollutants time series, the application of linear (MLR) and non–linear (NN) models to each monitoring station was performed. Linear MLR and non–linear NN models designed to forecast daily average PM_{10} concentrations in Lisbon, Portugal, were used to produce forecasts and hindcasts. The models were calibrated using air quality and meteorological data from 2002 until 2006 taken at 12 monitoring stations. The forecasting capabilities of the different approaches were then compared. The approach based on the removal of the weekly cycle presented the best results, comparatively to the use of the complete signal. Moreover, MLR and NN showed similar performances when evaluated by each of the above criteria. However, the TOT–MLR model had a significantly lower F and F50 false alarm rates. Therefore, we find it reasonable to conclude that there is no significant advantage on the use of NN against MLR for the case studied.

All in all, the models presented here are able to produce different results for each monitoring station, which allows a good spatial resolution for Lisbon's urban area. Consistent with the performance measures, high pollutant peak values were reproduced in most cases by each model. The simplicity and cost efficiency of these models, associated with their performance capabilities, show to be very promising for urban air quality characterization, allowing further developments in order to produce an integrated air quality surveillance system for the area of Lisbon. Being a general numerical procedure for any given set of measurements, our finding can be easily adapted to other NN models in weather or geophysical forecast.

Acknowledgments

The authors acknowledge the Instituto Português do Mar e da Atmosfera and the Agência Portuguesa do Ambiente for the meteorological and environmental data, respectively. The authors thank Deutscher Akademischer Auslandsdienst (DAAD) and Fundação para a Ciência e Tecnologia (FCT) through the bilateral cooperation DRI/DAAD/1208/2013 (FR and PGL) and FCT for support through SFRH/BPD/65427/2009 (FR). PL thanks BMU (German Environment Ministery) under the project 41V6451.

Supporting Material Available

Monthly distribution of pollutant's concentrations throughout the year for the entire studied period (2002–2006) and for all the monitoring stations (Figure S1a), Supplement to the box–plot analysis in Figure S1a just for Avenida da Liberdade (AL) monitoring station (Figure S1b), Accumulated number of PM_{10} exceeding values (PM_{10}>50 μg/m^3) for the period 2002–2006 (Figure S2), Scatter plots of MLR results versus actual observed PM_{10} values for each monitoring station and for the year 2006 (Figure S3). This information is available free of charge via the Internet at http://www.atmospolres.com.

List of abbreviations

AL: Avenida da Liberdade
ARIMA: Auto–regressive integrated moving average
BLH: Boundary layer height
BMU: German Environment Ministery
BP: Back–propagation
CMAQ: Community Multiscale Air Quality Model
CWT: Circulation Weather Types
DAAD: Deutscher Akademischer Auslandsdienst
E: Entrecampos
ECMWF: European Centre for Medium–Range Weather Forecasts
ERA–Interim: reanalysis of the global atmosphere covering the data–rich period since 1979
F: False alarm rate
F50: Legal limit value of false alarm rate
FCT: Fundacao para a Ciencia e Tecnologia
FSR: Forward stepwise regression
L: Lavradio
LM: Levenberg–Marquardt method
LMS: Least mean square error
MLR: Multiple linear regression
NN: Artificial Neural Networks
O: Olivais
PC: Pearson correlation coefficient
PCA: Principal component analysis
PCS: Proportion of correctness
PCS50: Proportion of correctness of predicted values above the legal limit value
PM_{10}: Particulate Matter, particularly, breathable particulate matter sized 10 μg or less
RMSE: Root mean square error
ROM: Regional Oxidant Model
SLP: Sea Level Pressure
SR: Stepwise regression
SSp: Skill against persistence
UAM: Urban Airshed Model
UTC: Universal Time Coordinated

References

Agirre–Basurko, E., Ibarra–Berastegi, G., Madariaga, I., 2006. Regression and multilayer perceptron–based models to forecast hourly O_3 and NO_2 levels in the Bilbao area. *Environmental Modelling & Software* 21, 430–446.

Almeida, S.M., Silva, A.I., Freitas, M.C., Dzung, H.M., Caseiro, A., Pio, C.A., 2013. Impact of Maritime Air Mass Trajectories on the Western European Coast Urban Aerosol, *Journal of Toxicology and Environmental Health, Part A* 76, 252- 262.

APA (Agencia Portuguesa do Ambiente), 2008. Evolution of the Portuguese Air Quality between the Years 2001 and 2005, 100 pages.

Arasa, R., Soler, M.R., Ortega, S., Olid, M., Merino, M., 2010. A performance evaluation of MM5/MNEQA/CMAQ air quality modelling system to forecast ozone concentrations in Catalonia. *Tethys* 7, 11 – 23.

Chaloulakou, A., Grivas, G., Spyrellis, N., 2003. Neural network and multiple regression models for PM_{10} prediction in Athens: A comparative assessment. *Journal of the Air & Waste Management Association* 53, 1183–1190.

Cobourn, W.G., Dolcine, L., French, M., Hubbard, M.C., 2000. A comparison of nonlinear regression and neural network models for ground–level ozone forecasting. *Journal of the Air & Waste Management Association* 50, 1999–2009.

Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. *Journal of the Air & Waste Management Association* 47, 653–663.

Dayan, U., Levy, I., 2005. The influence of meteorological conditions and atmospheric circulation types on PM_{10} and visibility in Tel Aviv. *Journal of Applied Meteorology* 44, 606–619.

Dayan, U., Levy, I., 2002. Relationship between synoptic–scale atmospheric circulation and ozone concentrations over Israel. *Journal of Geophysical Research–Atmospheres* 107, art. no. 4813.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge–Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.N., Vitart, F., 2011. The ERA–Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* 137, 553–597.

Demuzere, M., Trigo, R.M., Vila–Guerau de Arellano, J., van Lipzig, N.P.M., 2009. The impact of weather and atmospheric circulation on O_3 and PM_{10} levels at a rural mid–latitude site. *Atmospheric Chemistry and Physics* 9, 2695–2714.

Dias, D., Tchepel, O., Carvalho, A., Miranda, A.I., Borrego, C., 2012. Particulate matter and health risk under a changing climate: Assessment for Portugal. *Scientific World Journal*, art. no. 409546.

Diaz, J., Linares, C., Lopez, C., Garcia–Herrera, R., Trigo, R.M., 2004. Relationship between environmental factors and infant mortality in Madrid. *Journal of Occupational and Environmental Medicine* 6, 768–774.

Du, C., Liu, S., Yu, X., Li X., Chen, C., Peng, Y., Dong, Y., Dong, Z., Wang, F., 2013. Urban boundary layer height characteristics and relationship with particulate matter mass concentrations in Xi'an, Central China. *Aerosol and Air Quality Research* 13, 1598–1607.

Dutot, A.L., Rynkiewicz, J., Steiner, F.E., Rude, J., 2007. A 24–h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions. *Environmental Modelling & Software* 22, 1261–1269.

ECMWF (European Centre for Medium – Range Weather Forecasts), 2013. http://apps.ecmwf.int/datasets/, accessed in 2013.

EEA (European Environment Agency), 2013. Every Breath We Take: Improving Air Quality in Europe, EEA Signals 2013 Report, Copenhagen.

EEA (European Environment Agency), 2011. Air Quality in Europe — 2011 Report, EEA Technical Report, No 12/2011.

Elangasinghe, M.A., Singhal, N., Dirks, K.N., Salmond, J.A., 2014. Development of an ANN–based air pollution forecasting system with explicit knowledge through sensitivity analysis. *Atmospheric Pollution Research* 5, 696–708.

Garcia–Herrera, R., Paredes, D., Trigo, R.M., Trigo, I.F., Hernandez, E., Barriopedro, D., Mendes, M.A., 2007. The outstanding 2004/05 drought in the Iberian Peninsula: Associated atmospheric circulation. *Journal of Hydrometeorology* 8, 483–498.

Gardner, M.W., Dorling, S.R., 2000a. Meteorologically adjusted trends in UK daily maximum surface ozone concentrations. *Atmospheric Environment* 34, 171–176.

Gardner, M.W., Dorling, S.R., 2000b. Statistical surface ozone models: An improved methodology to account for non–linear behaviour. *Atmospheric Environment* 34, 21–34.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron) – A review of applications in the atmospheric sciences. *Atmospheric Environment* 32, 2627–2636.

Haykin, S., 1999. *Neural Networks – A Comprehensive Foundation*, College Publishing Company, New York.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., 2005. A neural network forecast for daily average PM_{10} concentrations in Belgium. *Atmospheric Environment* 39, 3279–3289.

Jolliffe I., Stephenson D., 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, John Wiley and Sons, Chichester, ISBN 0-471-49759-2.

Kolehmainen, M., Martikainen, H., Ruuskanen, J., 2001. Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment* 35, 815–825.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmospheric Environment* 37, 4539–4550.

Lal, B., Tripathy, S.S., 2012. Prediction of dust concentration in open cast coal mine using artificial neural network. *Atmospheric Pollution Research* 3, 211–218.

Luecken, D.J., Hutzell, W.T., Gipson, G.L., 2006. Development and analysis of air quality modeling simulations for hazardous air pollutants. *Atmospheric Environment* 40, 5087–5096.

Middleton, D.R., 1998. A new box model to forecast urban air quality: BOXURB. *Environmental Monitoring and Assessment* 52, 315–335.

Monteiro, A., Vautard, R., Borrego, C., Miranda, A.I., 2005. Long–term simulations of photo oxidant pollution over Portugal using the CHIMERE model. *Atmospheric Environment* 39, 3089–3101.

Nejadkoorki, F., Baroutian, S., 2012. Forecasting extreme PM_{10} concentrations using artificial neural networks. *International Journal of Environmental Research* 6, 277–284.

Papanastasiou, D.K., Melas, D., Kioutsioukis, I., 2007. Development and assessment of neural network and multiple regression models in order to predict PM_{10} levels in a medium–sized Mediterranean city. *Water Air and Soil Pollution* 182, 325–334.

Perez, P., Reyes, J., 2006. An integrated neural network model for PM_{10} forecasting. *Atmospheric Environment* 40, 2845–2851.

Russo, A., Trigo, R.M., Martins, H., Mendes, M.T., 2014. NO_2, PM_{10} and O_3 urban concentrations and its association with circulation weather types in Portugal. *Atmospheric Environment* 89, 768–785.

Russo, A., Raischel, F., Lind, P.G., 2013. Air quality prediction using optimal neural networks with stochastic variables. *Atmospheric Environment* 79, 822–830.

Saavedra, S., Rodriguez, A., Taboada, J.J., Souto, J.A., Casares, J.J., 2012. Synoptic patterns and air mass transport during ozone episodes in Northwestern Iberia. *Science of the Total Environment* 441, 97–110.

Shi, J.P., Harrison, R.M., 1997. Regression modelling of hourly NO_x and NO_2 concentrations in urban air in London. *Atmospheric Environment* 31, 4081–4094.

Stedman, J.R., King, K., Holland, M., Walton, H., 2002. Quantification of the Health Effects of Air Pollution in the UK for Revised PM_{10} Objective Analysis, DEFRA, AEAT/ENV/R/1162, Issue 1.

Trigo, R., DaCamara, C., 2000. Circulation weather types and their influence on the precipitation regime in Portugal. *International Journal of Climatology* 20, 1559–1581.

Trigo, R.M., Palutikof, J.P., 1999. Simulation of daily temperatures for climate change scenarios over Portugal: A neural network model approach. *Climate Research* 13, 45–59.

Trigo, R.M., Pereira, J.M.C., Pereira, M.G., Mota, B., Calado, T.J., Dacamara, C.C., Santo, F.E., 2006. Atmospheric conditions associated with the exceptional fire season of 2003 in Portugal. *International Journal of Climatology* 26, 1741–1757.

Weisberg, S., 1985. *Applied Linear Regression*, John Wiley & Sons, New York.

Wilks, D., 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd Edition, Academic Press, Burlington, MA.

Wong, C.M., Atkinson, R.W., Anderson, H.R., Hedley, A.J., Ma, S., Chau, P.Y.K., Lam, T.H., 2002. A tale of two cities: Effects of air pollution on hospital admissions in Hong Kong and London compared. *Environmental Health Perspectives* 110, 67–77.

Yi, J.S., Prybutok, V.R., 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution* 92, 349–357.